DTIC FILE COPY

(2)

AFHRL-TR-89-21

AD-A213 496

AIR FORCE HUMAN RESOURCES LABORATORY

PERFORMANCE MEASUREMENT DEVELOPMENT FOR AIR COMBAT

Gary S. Thomas

University of Dayton Research Institute
300 College Park Avenue
Dayton, Ohio 45469

David C. Miller

Logicon, Incorporated
Tactical and Training Systems Division
4010 Sorrento Valley Boulevard
San Diego, California 92138-5158

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85240-6457

October 1989

Interim Report for Period October 1987 - May 1989

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

89 10 10 048

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFHRL-TR-89-21

6a. NAME OF PERFORMING ORGANIZATION: University of Dayton Research Institute

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: Operations Training Division

6c. ADDRESS *(City, State, and ZIP Code)*: 300 College Park Avenue, Dayton, Ohio 45469

7b. ADDRESS *(City, State, and ZIP Code)*: Air Force Human Resources Laboratory, Williams Air Force Base, Arizona 85240-6457

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: Air Force Human Resources Laboratory

8b. OFFICE SYMBOL *(If applicable)*: HQ AFHRL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F33615-87-C-0012, F33615-86-C-0012

8c. ADDRESS *(City, State, and ZIP Code)*: Brooks Air Force Base, Texas 78235-5601

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
62205F	1123	35, 03	06, 83

11. TITLE *(Include Security Classification)*: Performance Measurement Development for Air Combat

12. PERSONAL AUTHOR(S): Thomas, G.S., Miller, D.C.

13a. TYPE OF REPORT: Interim

13b. TIME COVERED: FROM Oct 87 TO May 89

14. DATE OF REPORT *(Year, Month, Day)*: October 1989

15. PAGE COUNT: 26

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	08	/
12	04	/

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

air combat maneuvering, ~~F-15~~ JET AIRCRAFT, FIGHTER AIRC

air-to-air combat, F-16

expert judgment [illegible] performance measurement

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The purpose of this research was to formulate a unitary measure of performance for simulated one-versus-one, within visual range, air-to-air combat. The measure will serve as a criterion for the development and validation of specific measures of Air Combat Maneuvering (ACM) skill that can be used to provide diagnostic performance feedback to pilots. Three experiments were conducted in which fighter pilots served as judges and rank-ordered, from most to least desirable, hypothetical ACM engagement outcomes. Outcome variables included (a) whether or not the hypothetical pilot achieved a "kill"; (b) whether or not he survived the mission; (c) the percent of time the pilot was in an offensive, defensive, or neutral posture; (d) length of engagement; and (e) posture at the beginning and end of the engagement (offensive, defensive, or neutral). Inter-rater agreement among judges in Experiment I was high, with correlations ranging from .93 to .99. Pilots' rankings of engagement outcomes were subjected to linear regression analyses to derive equations that could be used as a unitary measure of ACM success. The regression equation in Experiment I accounted for 95% of the variance in rankings, and the composite regression model calculated in Experiment II accounted for more than 70% of the variance. Experiment III

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Nancy J. Allin, Chief, STINFO Branch

22b. TELEPHONE *(Include Area Code)*: (512) 536-3877

22c. OFFICE SYMBOL: AFHRL/SCV

DD Form 1473, JUN 86 — *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

19. (Concluded)

validated the performance model devised from data collected in the first two experiments. The performance model accounted for 99% of the variance in the predicted outcome data and accounted for 94% of the variance in data obtained from a separate group of fighter pilots.

SUMMARY

The purpose of this research was to formulate and validate a unitary or composite measure of performance for simulated one-versus-one, within visual range, air-to-air combat. This measure will serve as a criterion for the development and validation of diagnostic measures of air combat maneuvering (ACM) skill. The measure has utility for general pilot proficiency assessment, scaling the difficulty of training or practice exercises, evaluating the effectiveness of alternative training procedures, and providing general feedback to trainees. The work was accomplished through a joint effort by the Air Force Human Resources Laboratory, Logicon, Inc., and the University of Dayton Research Institute. Linear regression analyses were used to describe how mission-qualified F-15 and F-16 fighter pilots rank-ordered hypothetical ACM outcomes typical of those obtained on the Air Combat Maneuvering Instrumentation (ACMI) range and in the Simulator for Air-to-Air Combat (SAAC) at Luke Air Force Base, Arizona. Regression equations typically accounted for over 90% of the variance in pilots' rankings. The regression model of ACM performance developed in two experiments was validated and refined in a third experiment. The final performance model accounted for more than 99% of the variance in the predicted outcome data and accounted for 94% of the variance in data supplied by an independent group of fighter pilots.

PREFACE

This effort represents a portion of the training technology objectives as outlined in the AFHRL Research and Technology Plan. One of the goals is to identify and demonstrate training approaches that enhance the combat proficiency of tactical pilots. The work described in this report was conducted primarily under Work Unit 1123-3506, Validation and Refinement of Techniques for Air Combat Performance Assessment, Contract No. F33615-86-C-0012 by Logicon, Inc. Part of the work was accomplished jointly with the University of Dayton Research Institute under Work Unit 1123-0383 which is part of Contract No. F33615-87-C-0012, Flying Training Research Support. The effort developed and validated a composite measure of performance for simulated one-versus-one, within visual range, air-to-air combat.

The authors express appreciation to Ms. Suzanne Gular for her efforts in data collection, to Dr. Michael Houck for editorial comments and assistance in data analysis, and to Ms. Marge Keslin for manuscript preparation. Thanks also go to the F-15 and F-16 pilots of the 58th TTW and the 405th TTW at Luke AFB, Arizona, whose cooperation made the research possible.

TABLE OF CONTENTS

I. INTRODUCTION

A prerequisite to conducting air combat training research is the ability to measure relevant pilot performance while training air combat maneuvering (ACM). Different types of performance measures may be necessary depending upon the proposed use of the measures (Ciavarelli, 1987; Vreuls & Obermeyer, 1985). For training applications, valid measures of ACM performance are necessary for diagnostic training feedback and for pilot and training system evaluation. Although a variety of specific measures of ACM skills are desirable for diagnostic feedback, "composite" measures are important for a number of purposes, particularly in a training setting. Composite or unitary measures serve as general indicators for (a) assessing pilot proficiency, (b) scaling the difficulty of training or practice exercises, (c) evaluating the effectiveness of alternative training procedures, and (d) providing general feedback to trainees (Thomas, 1984). Unitary or composite measures are also potentially useful as criterion measures for the validation of specific measures of ACM skills.

Procedures for determining how multiple sources of information are combined into composite assessments can be found in the decision-making literature (see Slovic & Lichtenstein, 1971, for a review). One procedure, least-squares linear regression, has been used in a variety of research settings to describe how expert judges arrive at overall assessments of performance. It has been demonstrated that these linear representations describe the decision rules of judges; in fact, the regression model often has better predictive quality than do the judges themselves (e.g., Meehl, 1954, 1965). In particular, a "bootstrapping" technique, as reviewed by Dawes and Corrigan (1974), has been used to construct modeled representations of judges' decision rules. The validity of the model is then tested against a new set of values also assessed by judges. Alternatively, a linear model can be constructed to represent all judges and then compared to assessments of a different group of judges.

This procedure has been used successfully in developing and validating composite measures of mission accomplishment for the Army and Air Force. Thomas and Cochlin (1983) used a regression model to describe how expert Army judges combined measures of several objectives of a covering force/delay mission into a single measure of overall mission success. Hypothetical battle outcomes were used to develop the models of individual judges' decision rules. Individual regression models accounted for more than 95% of the variance in individual judges' ratings of the battle outcomes, and inter-rater agreement (Spearman's rho) ranged from .75 to .99. Validity of the predictive models was determined by comparing each judge's ratings of actual battle outcomes, generated by battalion command groups conducting covering force missions in a computer-driven battle simulation, to ratings of mission accomplishment predicted by the models. The individual models accounted for at least 94% of the variance in three of four judges' ratings of actual outcomes. When the individual models were combined into a single model, that model accounted for 92% of the variance in the average of the judges' ratings of actual outcomes. That model was found to be sensitive to training provided in Combined Arms Tactical Training Simulation (Thomas, Kaplan, & Barber, 1984).

Thomas (1984) used a similar procedure to develop a composite measure of performance for an A-10 close air support mission trained in a single cockpit simulator. Relevant dimensions of the mission were whether or not the pilot survived the mission, how many tanks and ground threats he destroyed, and whether or not he destroyed the enemy command post. Inter-rater agreement among the eight Air Force judges was high, as reflected in correlations among ratings that ranged from .68 to .99, with a median rho of .90. Judges' ratings were described using a single multiple regression equation. Validity of the model was established by comparing ratings predicted by the model to those obtained from a different group of judges who rated a different set of close air support mission outcomes. Correlations ranged from rho = .90 to rho = .99.

The purpose of the current research was to formulate a composite measure of performance for simulated one-versus-one, within visual range, air-to-air combat. In the first experiment, the relevant objectives of the mission were identified, and a multiple linear regression equation was formulated to describe judges' ratings of hypothetical engagement outcomes. The objective of Experiment II was to expand the least-squares regression equation formulated in Experiment I. Air Force pilots rated a set of mission outcomes used in the first experiment but with the inclusion of additional mission outcome variables. The third experiment validated a regression model formulated from data collected in the first two experiments.

II. EXPERIMENT I

Subjects

Subjects were nine F-15 and eight F-16 instructor pilots with more than 1,000 hours in the aircraft, plus three retired fighter pilots with more than 500 air-to-air combat hours. The instructor pilots were mission-qualified fighter pilots who were instructing in Replacement Training Units (RTUs) at Luke Air Force Base, Arizona.

Stimulus Materials

Stimulus materials were 112 (3- by 5-inch) index cards printed with the following hypothetical outcomes of air-to-air combat: (a) whether or not the pilot survived his mission; (b) whether or not the pilot achieved a "kill" against his opponent; (c) the percentage of engagement time the pilot spent in offensive, defensive, and neutral position with respect to his opponents (0%, 20%, 40%, 60%, or 80%); and (d) the total engagement time (1 or 5 minutes). These variables were selected based upon interviews with several mission-qualified Air Force pilots and pilots with prior air-to-air combat experience. For pilots asked to state the objectives of one-versus-one, air-to-air combat, the typical responses were to kill the enemy while avoid being killed, and to maximize offensive position while minimizing defensive posture. Stimulus cards included all possible realistic combinations of the levels of variables stated above, assuming a neutral start for opposing aircraft.

Procedures

Judges were informed that the purpose of the experiment was to develop a single measure of air-to-air combat mission accomplishment based on the variables included on the stimulus cards. The judges ranked the 112 combat outcomes from best to worst in terms of how well the hypothetical air-to-air combat missions were accomplished. Judges were also asked to suggest other mission variables that may be useful in describing air combat performance, and pilots were provided with general definitions of offensive, defensive, and neutral position (Appendix A).

Results

The rank-orders of engagement outcomes obtained from each judge were compared to determine the degree of agreement among judges. Spearman rho correlations, which were calculated with the rank-orders, ranged from .93 to .99, indicating that there was very high agreement among the judges. No apparent differences in rankings existed among F-15, F-16, and retired combat pilots.

Judges' rankings of the air-to-air mission outcomes were subjected to least-squares multivariate regression analyses to derive equations that describe how each judge rank-ordered the engagement outcomes. The regression equations accounted for high proportions of the variance in the data, by considering only linear components of main effects, as indicated by values of R^2 which ranged from .90 to .99, with a median R^2 of .96. The regression equations successfully described how judges ranked the engagement outcomes, and the equations reflected some similarities and differences in the relative importance of the engagement variables as judged by the pilots.

All judges ranked all outcomes where the pilot survived higher than outcomes where the pilot did not, and it was judged better to achieve a kill than not to achieve a kill. The majority of judges considered short engagements preferable if a kill was achieved, and longer engagements preferable if the pilot did not survive his mission. Some judges considered length of engagement to be of little consequence. Most judges thought it was more important to minimize defensive time than to maximize offensive time, but some judges considered more offensive time preferable to minimizing defensive time.

The data from all judges were combined and analyzed by least-squares linear regression. The resulting equation was: y = 68.50 + 53.60 (survival) + 27.13 (kill) -1.32 (engagement time) - .14 (% defensive) + .09 (% offensive). This equation accounted for 95% of the variance in the raw data, with the majority of variance accounted for by the survival and kill variables.

Mission accomplishment scores for each of the original 112 combat outcomes were calculated by substituting the values assigned to each variable into the composite equation and multiplying by the appropriate coefficients. The resultant scores were then rank-ordered and correlated with the rank-orders from the judges in order to determine the degree to which the composite model predicted judges' responses. Correlations between judges' rank-orders of mission outcomes and those derived from the composite model are quite high, ranging from rho = .94 to rho = .99. These very high correlations demonstrate that the composite model does very well at representing judges' assessments of air-to-air combat mission performance.

Discussion

The least-squares regression model for air-to-air combat performance developed in this experiment appears to be successful in describing how fighter pilots combined multiple sources of combat information to determine overall mission performance. Even though there was not total agreement among the judges as to the relative importance of some of the components of air-to-air mission performance, the high correlations between predictions of the composite model and the assessments of the judges indicate that the model accurately represents the judges' decision rules.

Although the results are encouraging in terms of mathematically representing decision rules applied to combat outcomes by expert judges, the model is limited because it applies only to air combat where the initial position of opponents is neutral. Experiment II extends the performance model by varying the starting and ending positions of the hypothetical engagements.

III. EXPERIMENT II

Based upon an analysis of performance measurement requirements and additional air combat variables suggested by the pilots in Experiment I, two additional variables were considered in Experiment II. The hypothetical engagements included an offensive, defensive, or neutral start of the engagement and an offensive, defensive, or neutral end of engagement. The hypothetical pilot survived his mission but did not kill the opponent in any engagement outcomes.

Subjects

Subjects were 12 F-16 (RTU) instructor pilots from Luke Air Force Base, Arizona. The pilots were mission-qualified fighter pilots with more than 1,000 hours in the aircraft who had not participated in Experiment I.

Stimulus Materials

Stimulus materials were 108 index cards (3- by 5-inch) printed with the following hypothetical outcomes of air-to-air combat: (a) the initial position of the hypothetical pilot with respect to his opponent (offensive, defensive, or neutral); (b) the position of the hypothetical pilot with respect to his opponent at the conclusion of the engagement (offensive, defensive, or neutral); (c) the percentage of engagement time the pilot spent in offensive, defensive, and neutral position with respect to his opponent (20%, 40%, 60%); and (d) the total engagement time (1 or 3 minutes). Experiment I pilots considered a 3-minute engagement more realistic than a 5-minute engagement.

Procedure

Pilots were informed that the purpose of the experiment was to develop a single measure of air-to-air combat mission accomplishment based on the variables included on the stimulus cards. Definitions of offensive, defensive, and neutral posture (Appendix A) and of beginning and end of the engagement were provided in the experiment instructions. These judges ranked the 108 combat outcomes from best to worst in terms of how well the hypothetical air-to-air combat missions were accomplished.

Results

Judges' rankings of engagement outcomes were correlated to determine inter-rater agreement. All correlations were significant beyond $p < .01$, and they ranged from .32 to .96. A median rho of .84 indicated relatively high agreement among pilots. Individual judges' rankings were analyzed by least-squares linear regression to derive equations that describe how the engagements were rank-ordered. Again the regression equations accounted for a high proportion of the variance in the data, as indicated by R^2 values that ranged from .70 to .96, with a median R^2 of .94. As in Experiment I, the regression equations were successful in describing how judges ranked the engagement outcomes, and the equations reflected similarities and some differences in the relative importance of the engagement variables.

Typically, it was desirable to end the engagement in an offensive position and to improve tactical position from start to end of the engagement (e.g., to start defensive and then end offensive). In general, ending position was considered more important than position change from start to end of the engagements. For example, seven of the twelve judges rated outcomes that started and ended in an offensive posture (no change in tactical position, but good end position) higher than outcomes where hypothetical pilots improved their positions from defensive to neutral. Similarly, two-thirds of the pilots rated outcomes higher if engagements began in an offensive position and ended neutral (loss of tactical position) than if engagements started and ended in a defensive position (no change in tactical position, but poor end position). As noted in Experiment I, the pilots indicated that it was desirable to minimize time spent defensively while maximizing offensive time.

The combined data were subjected to a linear regression analysis to derive a composite measure of mission accomplishment. The resulting equation, $y = 89.6 + 29.8$ (End Position) -1.5 (Engagement Time) $+ .20$ (% Offensive) $- .29$ (% Defensive) $- 9.3$ (Beginning Position), accounted for 71% of the variance in the raw data by considering only linear components of main effects. The regression equation successfully described how expert judges rated the mission outcomes, but the variance accounted for by the equation was not as high as that observed in Experiment I. This is likely due to the fact that subject agreement in judgments was not as high in this experiment as it was in the prior experiment. Probably agreement was higher for survival and kill variables used in Experiment I than for the start and end position variables used in Experiment II.

Discussion

As in Experiment I, it appears that the regression models developed in Experiment II adequately described both individual and combined judges' rankings of combat outcomes. Although lower in Experiment II, there was relatively high agreement among judges in terms of the ranks assigned to the engagement outcomes. Therefore, the least-squares regression models for air-to-air combat mission performance developed in this effort appear successful in mathematically describing how expert judges combined multiple sources of combat information to determine overall mission performance. What remains to be accomplished is a combination of the regression equations formulated in the two experiments into a more comprehensive performance measurement model.

IV. EXPERIMENT III

The purpose of this experiment was to formulate an overall or composite air combat performance model and to validate it with data collected from a different group of pilots. Based on regression analyses performed in Experiments I and II, a predicted rank-order of engagement outcomes was established. These outcomes included all the variable combinations used in each of the prior experiments. The rank-order of outcomes was described by a linear regression analysis and compared to judgments of a separate group of F-15 pilots.

Subjects

Subjects were six F-15 RTU instructor pilots at Luke Air Force Base, Arizona. Pilots were mission-qualified fighter pilots who have more than 1100 hours in the aircraft and who had not participated in Experiment I or II.

Stimulus Materials

Stimulus materials were 216 index cards (3- by 5-inch) printed with the following hypothetical outcomes of air-to-air combat: (a) whether or not the hypothetical pilot survived his mission; (b) whether or not the pilot achieved a "kill" against his opponent; (c) the percentage of engagement time the pilot spent in offensive, defensive, and neutral positions with respect to his opponent (20%, 40%, or 60%); (d) the total engagement time (1 or 3 minutes); (e) the position of the pilot with respect to his opponent at the beginning of the engagement (offensive, defensive, or neutral); and (f) the relative position of the pilot at the conclusion of the engagement (offensive, defensive, or neutral). The set of stimulus materials (n = 216) was ordered from best to worst based on previous data analyses. Cards were selected sequentially and distributed into three sets. Cards 1, 4, 7, 10, etc. were assigned to set one; cards 2, 5, 8, 11, etc. were placed in the second set; and cards 3, 6, 9, 12, etc. were assigned to the third set. The resulting three sets of 72 cards were duplicated to obtain six sets of outcomes. This procedure was used so that all samples were representative of the total set, so that each outcome was judged an equal number of times, and so that two pilots could judge each sample set.

Procedure

Pilots were informed that the purpose of the experiment was to develop a single measure of air-to-air combat mission accomplishment based on the variables included on the stimulus cards. Definitions of offensive, defensive, and neutral posture (Appendix A) and of engagement start and end were provided in the experiment instructions. Each judge rank-ordered one of the three sets of 72 combat outcomes from best to worst in terms of how well the hypothetical air-to-air combat missions were accomplished. The first two pilots ranked the same set of outcomes, the second two pilots ranked another set of outcomes, and the third pair of pilots ranked the last set of outcomes.

Results

Judges' rankings of engagement outcomes were correlated to determine inter-rater agreement. The correlation between the first two pilots' judgments was rho = .98, for the second two pilots the rho was = .89, and a rho = .93 was obtained for the third two pilots. These correlations indicate high agreement between the rank-orderings provided by the pilots. These rank-orders were also correlated with the predicted rank-order derived from Experiments I and II data analyses. As indicated below, obtained rank-orders compared very favorably to those predicted:

S_1 rho = .97, S_2 rho = .97, S_3 rho = .98,
S_4 rho = .97, S_5 rho = .96, S_6 rho = .98

The predicted rank-order of engagement outcomes was then subjected to a least-squares linear regression analysis to obtain a performance scoring algorithm. The resulting performance model, y = 217.2 + 106.5 (survival) + 20.4 (kill) -12 (start position) + 26.6 (end position) -.16 (% defensive) +.10 (% offensive) -.33 (time), accounted for 94% of the variance in the predicted outcome data.

Refined Performance Scoring Formula

Air combat performance scores were calculated by multiplying the coefficients of the linear regression formula by values assigned to levels of each variable. Scores ranged from 199.21 to 414.37. It was observed that scores for four combat outcomes where the pilot did not survive his mission were higher than some scores for outcomes where he did survive. This relationship was never observed in the rank-order data obtained in these experiments. Furthermore, scores for 15 outcomes where the pilot survived his mission but did not achieve a kill were higher than scores for outcomes where the pilot both survived his mission and achieved a kill. This relationship was rarely observed in the data. It was also observed that the regression formula placed slightly too much emphasis on starting position and too little emphasis on end position.

To rectify these minor anomalies in the regression equation, several coefficients were modified so that the regression model more closely matched the predicted data. The coefficients for survival, kill, and end position were increased, and the coefficient for start position was decreased. The constant 217.2 was reduced to 10, so that possible scores would range from about 0 to 245.

The regression model of air combat performance was refined and simplified for use as a performance scoring algorithm. The revised algorithm, y = 10 + 120 (survival) + 30 (kill) - 10 (start position) +.08 (% offensive) - .12 (% defensive) + 30 (end position) - .33 (time), can be used to score ACM performance by substituting empirically obtained values into the equation and multiplying by the appropriate coefficients. For example, if the pilot survives his mission, does not achieve a kill, begins and ends the engagement in a neutral position, and spends 60% of the 3-minute engagement in the offensive position and 20% in the defensive position, he would achieve a score of 171.41 out of a possible 245 points (see Appendix B). Scoring requires a value of 1 for survival and for a kill, 0 for nonsurvival and for no kill, 1 for defensive start and for defensive end position, 2 for neutral start and for neutral end position, 3 for offensive start and for offensive end position, and the absolute value obtained for the total time of engagement. Therefore, a score of 171.41 is obtained by the following calculation: 10 + 120 (1) + 30 (0) - 10 (2) +.08 (60) -.12 (20) + 30 (2) - .33 (3).

Appendix B lists the stimulus materials used in Experiment III, along with the corresponding performance scores calculated as shown above. As indicated in the appendix, there is a near-perfect degree of relationship (rho = .9999) between the rank-order of performance scores and the predicted rank-order of engagement outcomes. The relationship would be a perfect rho = 1.0 if the sign on the coefficient for "total time of engagement" is changed when the pilot does not survive and does not obtain a kill. The degree of relationship reflected in the Appendix is actually slightly higher than the correlation between predicted rank and rank of scores calculated by the least-squares linear regression model (rho = .9977). The final performance model is easier to implement for scoring air combat performance and actually describes more of the variance in the data than does the least-squares regression model.

Discussion

As observed in Experiment I, there was high agreement among pilots in terms of the rank-orders they assigned to the combat outcomes. Also replicating prior results, the linear regression equation formulated in this experiment accounted for a very large part of the variance in the predicted rank-order data. Further, the data described by the air combat performance model compare very favorably to data obtained in Experiment III. The fact that the performance algorithm predicts the data collected in this experiment indicates that the model accurately represents the air combat performance judgments of pilots. This process can be considered a partial validation of the performance algorithm developed in this experiment.

V. CONCLUSION

The ACM performance algorithm represents an advance over simple loss exchange ratios and is more precise in that outcomes where no kill is achieved still result in performance scores. The performance model can also be implemented as a machine-scorable algorithm to evaluate ACM performance when expert pilots are not available to judge ACM outcomes.

The model can be used in research to validate other more specific measures of ACM skill that can be used for diagnostic training feedback. Such measures include the All-Aspect Maneuvering Index (McGuinness, Forbes, & Rhoads, 1984), an energy management index (Pruitt, Moroney, & Lau, 1980), and discriminant measures of ACM performance (Kelly et al., 1979). This research could also provide further validation of the measurement model as recommended by Lane (1986) and Waag and Knoop (1977).

The performance measurement algorithm developed in this research has potential for use as a general indicator of pilot proficiency in ACM before and after ACM training to assess the effectiveness of alternative training methods and devices. The methodology of developing a unitary measure of ACM performance using linear regression analysis may also be applicable to more complicated multiship scenarios such as sweep, force protection, and point defense missions. This could be accomplished by quantifying the objectives of these missions and by using procedures similar to those reported here to develop performance measures for each mission.

REFERENCES

Ciavarelli, A.P. (1987). Development and validation of aircrew performance measures. Paper presented at the Department of Defense/Educational Testing Service Conference on Job Performance Measurement Technologies, San Diego, CA.

Dawes, R.M., & Corrigan, B. (1974). Linear models in decision making. Psychological Bulletin, 81, 95-106.

Kelly, M.J., Wooldridge, L., Hennessy, R.T., Vreuls, D., Barnebey, S.F.,Cotton, J.C., & Reed, J.C. (1979). Air combat maneuvering performance measurement (NAVTRAEQUIPCEN IH-35/AFHRL-TR-79-3, AD-A077 429). Orlando, FL: Naval Training Equipment Center, and Williams Air Force Base, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Lane, N.E. (1986). Issues in performance measurement for military aviation with applications to air combat maneuvering. Orlando, FL: Naval Training Systems Center: Center of Excellence for Simulation and Training Technology.

McGuinness, J., Forbes, J.M., & Rhoads, J.E. (1984). Air combat maneuvering performance measurement system design (AFHRL-TP-83-56, AD-B081 060L). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Meehl, P.E. (1954). Clinical versus statistical prediction: A theoretical analysis and review of the literature. Minneapolis, MN: University of Minnesota Press.

Meehl, P.E. (1965). Seer over sign: The first good example. Journal of Experimental Research in Personality, 1, 27-32.

Pruitt, V.R., Moroney, W.F., & Lau, C. (1980). Energy maneuverability display for the air combat maneuvering range/tactical training system (ACMR/TACTS) (Special Report 80-4). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.

Thomas, G.S. (1984). Close air support mission: development of a unitary measure of pilot performance (AFHRL-TR-84-39, AD-A149 285). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Thomas, G.S., & Cochlin, T.G. (1983). A model of mission accomplishment in simulated battle (ARI Technical Report 599). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Thomas, G.S., Kaplan, I.T., & Barber, H.F. (1984). Command and control training in the combined arms tactical training simulator (ARI Technical Report 615). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Vreuls, D., & Obermeyer, R.W. (1985). Human-system performance measurement in training simulators. Human Factors, 27, 241-250.

Waag, W.L., & Knoop, P.A. (1977). Planning for aircrew performance productivity enhancement R&D: U.S. Air Force. In L.T. Pope & D. Meister (Eds.), Productivity enhancement: Personnel performance assessment in Navy systems. San Diego, CA: Navy Personnel Research and Development Center.

APPENDIX A: INSTRUCTIONS

INSTRUCTIONS

The purpose of this study is to develop a unitary outcome measure of ACM performance based on expert judgments taken from individuals such as yourself. The outcome measure will discriminate between "winners" and "losers" in 1v1, within visual range engagements, and be used in later research to develop other performance measures for training feedback.

Your task is to rank-order from best to worst a set of hypothetical ACM outcomes achieved by one pilot. Outcome variables include whether or not the pilot achieves a kill, whether or not he survives the engagement, the percent of engagement time he spends offensive, defensive, or neutral, and total engagement time. In the sorting task, assume similar aircraft and a neutral start as you sort the outcomes from best to worst.

Definitions of Indices:

"Kill/No Kill" refers to whether you have killed or not killed your opponent by the end of the engagement.

"Survived/No Survive" refers to whether or not you have survived the engagement.

"% Time Offensive" refers to what percent of the entire engagement you were in an offensive tactical position; i.e., in a position where the opponent must maneuver to avoid your shooting at him, or when you are in a position to force the fight.

"% Time Defensive" refers to what percent of the entire engagement you were in a defensive tactical position, a position where the opponent is forcing the fight; i.e., you must avoid being shot at by your opponent.

"% Time Neutral" refers to what percent of the entire engagement you were in a neutral tactical position; i.e., in a position where you cannot bring your weapons to bear on your opponent, and he also cannot bring his weapons to bear on you.

"Length of Engagement" refers to how long the entire engagement lasts.

APPENDIX B: PERFORMANCE SCORES AND PREDICTED RANK-ORDER OF EXPERIMENT III COMBAT OUTCOMES

Predicted Rank	Survive	Kill	Start Position	% Time Offensive	% Time Defensive	% Time Neutral	End Position	Time of Engagement	Stimulus Number	Stimulus Set	Performance Score
1	1	1	1	60	20	20	3	1	1	1	242.07
2	1	1	1	60	20	20	3	3	7	2	241.41
3	1	1	1	40	20	40	3	1	2	3	240.47
4	1	1	1	40	20	40	3	3	8	1	239.81
5	1	1	1	20	20	60	3	1	3	2	238.87
6	1	1	1	20	20	60	3	3	9	3	238.21
7	1	1	1	40	40	20	3	1	4	1	238.07
8	1	1	1	40	40	20	3	3	10	2	237.41
9	1	1	1	20	40	40	3	1	5	3	236.47
10	1	1	1	20	40	40	3	3	11	1	235.81
11	1	1	1	20	60	20	3	1	6	2	234.07
12	1	1	1	20	60	20	3	3	12	3	233.41
13	1	1	2	60	20	20	3	1	13	1	232.07
14	1	1	2	60	20	20	3	3	19	2	231.41
15	1	1	2	40	20	40	3	1	14	3	230.47
16	1	1	2	40	20	40	3	3	20	1	229.81
17	1	1	2	20	20	60	3	1	15	2	228.87
18	1	1	2	20	20	60	3	3	21	3	228.21
19	1	1	2	40	40	20	3	1	16	1	228.07
20	1	1	2	40	40	20	3	3	22	2	227.41
21	1	1	2	20	40	40	3	1	17	3	226.47
22	1	1	2	20	40	40	3	3	23	1	225.81
23	1	1	2	20	60	20	3	1	18	2	224.07
24	1	1	2	20	60	20	3	3	24	3	223.41
25	1	1	3	60	20	20	3	1	25	1	222.07
26	1	1	3	60	20	20	3	3	31	2	221.41
27	1	1	3	40	20	40	3	1	26	3	220.47
28	1	1	3	40	20	40	3	3	32	1	219.81
29	1	1	3	20	20	60	3	1	27	2	218.87
30	1	1	3	20	20	60	3	3	33	3	218.21
31	1	1	3	40	40	20	3	1	28	1	218.07
32	1	1	3	40	40	20	3	3	34	2	217.41
33	1	1	3	20	40	40	3	1	29	3	216.47
34	1	1	3	20	40	40	3	3	35	1	215.81
35	1	1	3	20	60	20	3	1	30	2	214.07
36	1	1	3	20	60	20	3	3	36	3	213.41
37	1	0	1	60	20	20	3	1	37	1	212.07
38	1	0	1	60	20	20	3	3	43	2	211.41
39	1	0	1	40	20	40	3	1	38	3	210.47
40	1	0	1	40	20	40	3	3	44	1	209.81
41	1	0	1	20	20	60	3	1	39	2	208.87
42	1	0	1	20	20	60	3	3	45	3	208.21
43	1	0	1	40	40	20	3	1	40	1	208.07
44	1	0	1	40	40	20	3	3	46	2	207.41
45	1	0	1	20	40	40	3	1	41	3	206.47
46	1	0	1	20	40	40	3	3	47	1	205.81
47	1	0	1	20	60	20	3	1	42	2	204.07
48	1	0	1	20	60	20	3	3	48	3	203.41
49	1	0	2	60	20	20	3	1	49	1	202.07
50	1	0	2	60	20	20	3	3	55	2	201.41

51	1	0	2	40	20	40	3	1	50	3	200.47
52	1	0	2	40	20	40	3	3	56	1	199.81
53	1	0	2	20	20	60	3	1	51	2	198.87
54	1	0	2	20	20	60	3	3	57	3	198.21
55	1	0	2	40	40	20	3	1	52	1	198.07
56	1	0	2	40	40	20	3	3	58	2	197.41
57	1	0	2	20	40	40	3	1	53	3	196.47
58	1	0	2	20	40	40	3	3	59	1	195.81
59	1	0	2	20	60	20	3	1	54	2	194.07
60	1	0	2	20	60	20	3	3	60	3	193.41
61	1	0	3	60	20	20	3	1	61	1	192.07
62	1	0	3	60	20	20	3	3	67	2	191.41
63	1	0	3	40	20	40	3	1	62	3	190.47
64	1	0	3	40	20	40	3	3	68	1	189.81
65	1	0	3	20	20	60	3	1	63	2	188.87
66	1	0	3	20	20	60	3	3	69	3	188.21
67	1	0	3	40	40	20	3	1	64	1	188.07
68	1	0	3	40	40	20	3	3	70	2	187.41
69	1	0	3	20	40	40	3	1	65	3	186.47
70	1	0	3	[illegible]0	40	40	3	3	71	1	185.81
71	1	0	3	20	60	20	3	1	66	2	184.07
72	1	0	3	20	60	20	3	3	72	3	183.41
73	1	0	1	60	20	20	2	1	73	1	182.07
74	1	0	1	60	20	20	2	3	79	2	181.41
75	1	0	1	40	20	40	2	1	74	3	180.47
76	1	0	1	40	20	40	2	3	80	1	179.81
77	1	0	1	20	20	60	2	1	75	2	178.87
78	1	0	1	20	20	60	2	3	81	3	178.21
79	1	0	1	40	40	20	2	1	76	1	178.07
80	1	0	1	40	40	20	2	3	82	2	177.41
81	1	0	1	20	40	40	2	1	77	3	176.47
82	1	0	1	20	40	40	2	3	83	1	175.81
83	1	0	1	20	60	20	2	1	78	2	174.07
84	1	0	1	20	60	20	2	3	84	3	173.41
85	1	0	2	60	20	20	2	1	85	1	172.07
86	1	0	2	60	20	20	2	3	91	2	171.41
87	1	0	2	40	20	40	2	1	86	3	170.47
88	1	0	2	40	20	40	2	3	92	1	169.81
89	1	0	2	20	20	60	2	1	87	2	168.87
90	1	0	2	20	20	60	2	3	93	3	168.21
91	1	0	2	40	40	20	2	1	88	1	168.07
92	1	0	2	40	40	20	2	3	94	2	167.41
93	1	0	2	20	40	40	2	1	89	3	166.47
94	1	0	2	20	40	40	2	3	95	1	165.81
95	1	0	2	20	60	20	2	1	90	2	164.07
96	1	0	2	20	60	20	2	3	96	3	163.41
97	1	0	3	60	20	20	2	1	97	1	162.07
98	1	0	3	60	20	20	2	3	103	2	161.41
99	1	0	3	40	20	40	2	1	98	3	160.47
100	1	0	3	40	20	40	2	3	104	1	159.81
101	1	0	3	20	20	60	2	1	99	2	158.87
102	1	0	3	20	20	60	2	3	105	3	158.21
103	1	0	3	40	40	20	2	1	100	1	158.07
104	1	0	3	40	40	20	2	3	106	2	157.41
105	1	0	3	20	40	40	2	1	101	3	156.47
106	1	0	3	20	40	40	2	3	107	1	155.81

107	1	0	3	20	60	20	2	1	102	2	154.07
108	1	0	3	20	60	20	2	3	108	3	153.41
109	1	0	1	60	20	20	1	1	109	1	152.07
110	1	0	1	60	20	20	1	3	115	2	151.41
111	1	0	1	40	20	40	1	1	110	3	150.47
112	1	0	1	40	20	40	1	3	116	1	149.81
113	1	0	1	20	20	60	1	1	111	2	148.87
114	1	0	1	20	20	60	1	3	117	3	148.21
115	1	0	1	40	40	20	1	1	112	1	148.07
116	1	0	1	40	40	20	1	3	118	2	147.41
117	1	0	1	20	40	40	1	1	113	3	146.47
118	1	0	1	20	40	40	1	3	119	1	145.81
119	1	0	1	20	60	20	1	1	114	2	144.07
120	1	0	1	20	60	20	1	3	120	3	143.41
121	1	0	2	60	20	20	1	1	121	1	142.07
122	1	0	2	60	20	20	1	3	127	2	141.41
123	1	0	2	40	20	40	1	1	122	3	140.47
124	1	0	2	40	20	40	1	3	128	1	139.81
125	1	0	2	20	20	60	1	1	123	2	138.87
126	1	0	2	20	20	60	1	3	129	3	138.21
127	1	0	2	40	40	20	1	1	124	1	138.07
128	1	0	2	40	40	20	1	3	130	2	137.41
129	1	0	2	20	40	40	1	1	125	3	136.47
130	1	0	2	20	40	40	1	3	131	1	135.81
131	1	0	2	20	60	20	1	1	126	2	134.07
132	1	0	2	20	60	20	1	3	132	3	133.41
133	1	0	3	60	20	20	1	1	133	1	132.07
134	1	0	3	60	20	20	1	3	139	2	131.41
135	1	0	3	40	20	40	1	1	134	3	130.47
136	1	0	3	40	20	40	1	3	140	1	129.81
137	1	0	3	20	20	60	1	1	135	2	128.87
138	1	0	3	20	20	60	1	3	141	3	128.21
139	1	0	3	40	40	20	1	1	136	1	128.07
140	1	0	3	40	40	20	1	3	142	2	127.41
141	1	0	3	20	40	40	1	1	137	3	126.47
142	1	0	3	20	40	40	1	3	143	1	125.81
143	1	0	3	20	60	20	1	1	138	2	124.07
144	1	0	3	20	60	20	1	3	144	3	123.41
145	0	1	1	60	20	20	3	1	145	1	122.07
146	0	1	1	60	20	20	3	3	151	2	121.41
147	0	1	1	40	20	40	3	1	146	3	120.47
148	0	1	1	40	20	40	3	3	152	1	119.81
149	0	1	1	20	20	60	3	1	147	2	118.87
150	0	1	1	20	20	60	3	3	153	3	118.21
151	0	1	1	40	40	20	3	1	148	1	118.07
152	0	1	1	40	40	20	3	3	154	2	117.41
153	0	1	1	20	40	40	3	1	149	3	116.47
154	0	1	1	20	40	40	3	3	155	1	115.81
155	0	1	1	20	60	20	3	1	150	2	114.07
156	0	1	1	20	60	20	3	3	156	3	113.41
157	0	1	2	60	20	20	3	1	157	1	112.07
158	0	1	2	60	20	20	3	3	163	2	111.41
159	0	1	2	40	20	40	3	1	158	3	110.47
160	0	1	2	40	20	40	3	3	164	1	109.81
161	0	1	2	20	20	60	3	1	159	2	108.87
162	0	1	2	20	20	60	3	3	165	3	108.21

163	0	1	2	40	40	20	3	1	160	1	108.07
164	0	1	2	40	40	20	3	3	166	2	107.41
165	0	1	2	20	40	40	3	1	161	3	106.47
166	0	1	2	20	40	40	3	3	167	1	105.81
167	0	1	2	20	60	20	3	1	162	2	104.07
168	0	1	2	20	60	20	3	3	168	3	103.41
169	0	1	3	60	20	20	3	1	169	1	102.07
170	0	1	3	60	20	20	3	3	175	2	101.41
171	0	1	3	40	20	40	3	1	170	3	100.47
172	0	1	3	40	20	40	3	3	176	1	99.81
173	0	1	3	20	20	60	3	1	171	2	98.87
174	0	1	3	20	20	60	3	3	177	3	98.21
175	0	1	3	40	40	20	3	1	172	1	98.07
176	0	1	3	40	40	20	3	3	178	2	97.41
177	0	1	3	20	40	40	3	1	173	3	96.47
178	0	1	3	20	40	40	3	3	179	1	95.81
179	0	1	3	20	60	20	3	1	174	2	94.07
180	0	1	3	20	60	20	3	3	180	3	93.41
181	0	0	1	60	20	20	1	3	187	1	31.41
182	0	0	1	60	20	20	1	1	181	2	32.07
183	0	0	1	40	20	40	1	3	188	3	29.81
184	0	0	1	40	20	40	1	1	182	1	30.47
185	0	0	1	20	20	60	1	3	189	2	28.21
186	0	0	1	20	20	60	1	1	183	3	28.87
187	0	0	1	40	40	20	1	3	190	1	27.41
188	0	0	1	40	40	20	1	1	184	2	28.07
189	0	0	1	20	40	40	1	3	191	3	25.81
190	0	0	1	20	40	40	1	1	185	1	26.47
191	0	0	1	20	60	20	1	3	192	2	23.41
192	0	0	1	20	60	20	1	1	186	3	24.07
193	0	0	2	60	20	20	1	3	199	1	21.41
194	0	0	2	60	20	20	1	1	193	2	22.07
195	0	0	2	40	20	40	1	3	200	3	19.81
196	0	0	2	40	20	40	1	1	194	1	20.47
197	0	0	2	20	20	60	1	3	201	2	18.21
198	0	0	2	20	20	60	1	1	195	3	18.87
199	0	0	2	40	40	20	1	3	202	1	17.41
200	0	0	2	40	40	20	1	1	196	2	18.07
201	0	0	2	20	40	40	1	3	203	3	15.81
202	0	0	2	20	40	40	1	1	197	1	16.47
203	0	0	2	20	60	20	1	3	204	2	13.41
204	0	0	2	20	60	20	1	1	198	3	14.07
205	0	0	3	60	20	20	1	3	211	1	11.41
206	0	0	3	60	20	20	1	1	205	2	12.07
207	0	0	3	40	20	40	1	3	212	3	9.81
208	0	0	3	40	20	40	1	1	206	1	10.47
209	0	0	3	20	20	60	1	3	213	2	8.21
210	0	0	3	20	20	60	1	1	207	3	8.87
211	0	0	3	40	40	20	1	3	214	1	7.41
212	0	0	3	40	40	20	1	1	208	2	8.07
213	0	0	3	20	40	40	1	3	215	3	5.81
214	0	0	3	20	40	40	1	1	209	1	6.47
215	0	0	3	20	60	20	1	3	216	2	3.41
216	0	0	3	20	60	20	1	1	210	3	4.07
0	0	0	3	10	80	10	1	3			0.21
0	1	1	1	80	10	10	3	1			244.87

☆ U S GOVERNMENT PRINTING OFFICE 1989—661-041/9-00058